

Russia2019_14

Supporting the regular georeferencing of European Russia holdings in the Moscow Digital Herbarium (<https://plant.depo.msu.ru/>)

MID-TERM ACTIVITY REPORT

Contents

Executive summary	1
Contact information	2
Introduction	2
The project and its objectives	2
Project activities completed by mid-term.....	4
Project communications	6
Mid-term evaluation findings and recommendations for the remaining project implementation period	6
Annex – Sources of verification	15

Executive summary

The Moscow Digital Herbarium is the largest biodiversity database in Russia and at present the largest Russian dataset published in GBIF, but only 31% of the records were georeferenced by December, 30 2019. The main objective of the project is to increase the amount of geodata associated with these records. The total number of geodata from European Russia and Northern Caucasus at the starting point of the project (March, 1 2019) was 121,690 georeferences. We newly georeferenced 36,499 specimens completing the mission on 67% (out of 54,450). Twelve larger spreadsheets from various regions of European Russia included 24,684 new georeferences. Other 11,815 new georeferences were made by the herbarium staff members as their daily routine job and inserted through the web-interface. As of May, 15 2019, Moscow Digital Herbarium holds 158,189 georeferenced specimens from the focused area. Many georeferences were figured out

automatically using ISTR system due to permanent adding of new label transcriptions. In March–May 2019, we added to the system 119,962 fully captured labels from European Russia and Northern Caucasus. We are planning to focus further efforts on remaining spatial gaps present in Northern Region of European Russia, Volga-Kama Region and Lower Volga Region. We do not have much holdings from that regions, but we really need more progress in these areas. We are planning to finish georeferencing by mid-September 2019, after that to check 26,000 label transcriptions made by the commercial partner and to synchronize geodata and text data on the portal within one or two working days. The final report will be submitted to GBIF Secretariat by September, 30 2019.

Contact information

Name	Alexey P. Seregin, Dr. Sci.
Institution	M.V. Lomonosov Moscow State University
Role in project	Project leader

Introduction

The Moscow Digital Herbarium is the largest biodiversity database in Russia and at present the largest Russian dataset published in GBIF, but only 31% of the records were georeferenced by December, 30 2019. So, the main objective of the project is to increase the amount of geodata associated with these records. The report is highlighting some key quantitative indicators of the project progress with necessary explanations in two aspects. The first aspect is focused on a project timeline with a number of key dates when we assessed the current state of the deals. The second aspect is concerning a spatial coverage of the geodata by May, 15 2019 with a brief description of the spreadsheets integrated into the Moscow Digital Herbarium.

The analytical data for the report timeline were mined directly from <https://plant.depo.msu.ru/> at 01.03.2019 (starting point), 16.03.2019 (stage 1), 20.03.2019 (stage 2), 27.03.2019 (stage 3), 07.04.2019 (stage 4), 15.04.2019 (stage 5), 29.04.2019 (stage 6), and 15.05.2019 (mid-term report). The spatial data were analysed at 15.05.2019. Screenshots with maps were produced at 01.03.2019 (starting point) and 15.05.2019. The evaluation results are shown either as numbers of georeferenced records and in percentages to enable the partners and GBIF Secretariat to evaluate the project progress in a simple manner.

The project and its objectives

The project is aimed to intensify the georeferencing of European Russia holdings already imaged for the Moscow Digital Herbarium. We started the work in March, 01 2019 and aimed

to finish the mission by mid-September 2019. The final report will be delivered to GBIF secretariat by September, 30 2019.

The participants of the project are staff members of the Moscow University Herbarium, freelancers permanently involved in our activities, IT-staff of depo.msu.ru platform, and a commercial partner contracted for label capturing. The main stakeholder of the project is an international scientific community interested in filling geographical gaps in freely-accessible data relating to plant diversity of the European Russia. This is demonstrated by the growing volume of peer-reviewed research using data from our dataset discovered and accessed through GBIF.

So, the main objective of the project is to increase the volume of geodata within the largest GBIF-dataset published by the Russian institutions. The complete grant budget will be spent for text capturing of 26,000 labels by a commercial partner. That is essential to synchronize text data from the original labels with georeferencing activities implemented by the herbarium staff members and freelancers. This synchronization will give us ca. 5,000 automatic georeferences identified by the ISTR system. Initially, we planned to contract manual text capturing of 33,800 labels, but this figure was ceased after budget cuts. As a result, we pledged to georeference 54,450 specimens to reach the final number of 176,140 specimens with coordinates from the European Russia in the Moscow Digital Herbarium by 30.09.2019.

All results are being published on a regular basis as part of weekly synchronization of our database (available at <https://plant.depo.msu.ru/>) with the GBIF dataset (<https://doi.org/10.15468/cpnhcc>).

Project activities completed by mid-term

a. Data

Title of dataset	Taxonomic/ geographic/ temporal scope	Approximate number of records	Sampling methodology/protocol used (if relevant)	Geographic accuracy for most records (in m or km, or province, country etc.)	Current state (e.g. undigitized, digitized)	State by mid- term report
Moscow University Herbarium (MW)	Vascular plants; Global; 1739– 2018	975,000 records, including 54,450 newly georeferenced ones	Standard procedures of manual georeferencing using open-source web- platforms and map libraries; ISTRA algorithm for aggregation of identical labels	1–10 km	Herbarium specimens are imaged and published in GBIF; 14% got captured labels	36,499 new georeferences added out of 54,450 (67%); contract with a commercial partner for label capturing was discussed and approved.

Expected milestones by mid-term report

Milestone	Status by mid-term progress report
<p>At least one dataset has been published to GBIF.org</p>	<p>"Moscow University Herbarium (MW)" (https://doi.org/10.15468/cpnhcc) which includes 36,499 newly georeferenced specimens from European Russia and the following formal disclaimer in the metadata (Description of the dataset section):</p> <p>"36,499 occurrences were georeferenced from March, 1 2019 to May, 15 2019 by the support of FinBIF project 'Supporting the regular georeferencing of European Russia holdings in the Moscow Digital Herbarium' (Russia2019_14, see details at https://www.gbif.org/project/2dfnq4VJQxHVSOPZWykiCb/supporting-the-regular-georeferencing-of-european-russia-holdings-in-the-moscow-digital-herbarium). They include newly georeferenced records from Eastern European plain, western slope of the Urals and Northern Caucasus."</p>

Project communications

To deliver our experience in the collection digitization and data mobilization to a wider audience the project leader Dr. A.P. Seregin made three invited talks on the seminar “Natural sciences collections of Yugra: collecting, fixation, storage, delivering to a scientific community” (Khanty-Mansiysk, March 25-26, 2019). The links to videos are given in the Annex.

Dr. A.P. Seregin was also asked to make an invited lecture on 114th Società Botanica Italiana Congress which will take place from 4 to 7 September 2019 in Padova. The invitation was accepted and the talk entitled "Moscow Digital Herbarium: a global approach through the regional actions" will be given at the Symposium "Herbaria: still relevant in the 21th century?"

Some key points of the project are published in the Herbarium blog in the Russian social network VK at https://vk.com/mw_herbarium in the form of short posts. We are communicating with a wider audience this way and getting some feedback from the users. The links to some posts are given in the Annex.

Mid-term evaluation findings and recommendations for the remaining project implementation period

1. The project timeline

Here we present a number of key dates when we assessed the current state of the deals. The analytical data for the report timeline were mined directly from <https://plant.depo.msu.ru/> at 01.03.2019 (starting point), 16.03.2019 (stage 1), 20.03.2019 (stage 2), 27.03.2019 (stage 3), 07.04.2019 (stage 4), 15.04.2019 (stage 5), 29.04.2019 (stage 6), and 15.05.2019 (mid-term report).

Georeferences as of 01.03.2019 (starting point)

As of 01.03.2019, we imaged and published online 384,201 herbarium specimens from European Russia and Northern Caucasus, including 121,690 georeferenced specimens (31.6%).

Georeferences as of 16.03.2019 (stage 1)

European Russia, specimens	331,748
Northern Caucasus, specimens	55,037
TOTAL specimens	386,785
European Russia, georeferences	105,440
Northern Caucasus, georeferences	21,763
TOTAL georeferences	127,203
Georeferences added	5,513
Completion of the mission, %	10.1

Georeferences as of 20.03.2019 (stage 2)

European Russia, specimens	331,731
Northern Caucasus, specimens	55,039
TOTAL specimens	386,770
European Russia, georeferences	110,997
Northern Caucasus, georeferences	21,800
TOTAL georeferences	132,797
Georeferences added	11,107
Completion of the mission, %	20.4

Georeferences as of 27.03.2019 (stage 3)

European Russia, specimens	331,812
Northern Caucasus, specimens	55,035
TOTAL specimens	386,847
European Russia, georeferences	122,042
Northern Caucasus, georeferences	21,800
TOTAL georeferences	143,842
Georeferences added	22,152
Completion of the mission, %	40.7

Georeferences as of 07.04.2019 (stage 4)

European Russia, specimens	331,912
Northern Caucasus, specimens	55,035
TOTAL specimens	386,947

European Russia, georeferences	125,582
Northern Caucasus, georeferences	22,078
TOTAL georeferences	147,660
Georeferences added	25,970
Completion of the mission, %	47.7

Georeferences as of 15.04.2019 (stage 5)

European Russia, specimens	331,925
Northern Caucasus, specimens	55,035
TOTAL specimens	386,960
European Russia, georeferences	134,691
Northern Caucasus, georeferences	22,078
TOTAL georeferences	156,769
Georeferences added	35,079
Completion of the mission, %	64.4

Georeferences as of 29.04.2019 (stage 6)

European Russia, specimens	333,077
Northern Caucasus, specimens	55,035
TOTAL specimens	388,112
European Russia, georeferences	134,881
Northern Caucasus, georeferences	22,078
TOTAL georeferences	156,959
Georeferences added	35,269
Completion of the mission, %	64.8

Georeferences as of 15.05.2019 (mid-term report)

European Russia, specimens	333,077
Northern Caucasus, specimens	56,611
TOTAL specimens	389,688
European Russia, georeferences	135,260
Northern Caucasus, georeferences	22,929
TOTAL georeferences	158,189
Georeferences added	36,499

Completion of the mission, % 67.0

As of 15.05.2019, we imaged and published online 389,688 herbarium specimens from European Russia and Northern Caucasus, including 158,189 georeferenced specimens (40.6%).

Many georeferences were figured out automatically using ISTR system due to **permanent adding of new label transcriptions**. In March–May 2019, we added to the system 119,962 fully captured labels from European Russia (114,464) and Northern Caucasus (5,468). This work was made by the freelancers involved in our activities and regarded here as a co-funding of the project.

Label transcriptions as of 01.03.2019 (starting point)

European Russia, specimens	331,748
Northern Caucasus, specimens	55,037
TOTAL specimens	386,785
European Russia, transcriptions	37,349
Northern Caucasus, transcriptions	16,006
TOTAL transcriptions	53,355
Level of full-text capturing, %	13.8

Label transcriptions as of 15.05.2019 (mid-term report)

European Russia, specimens	333,077
Northern Caucasus, specimens	56,611
TOTAL specimens	389,688
European Russia, transcriptions	151,813
Northern Caucasus, transcriptions	21,504
TOTAL transcriptions	173,217
Transcriptions added	119,962
Level of full-text capturing, %	30.8

2. The project spatial coverage

The total number of geodata at the starting point of the project (March, 1 2019) was 121,690 georeferences (Fig. 1). As of May, 15 2019, we georeferenced 158,189 specimens (Fig. 2).

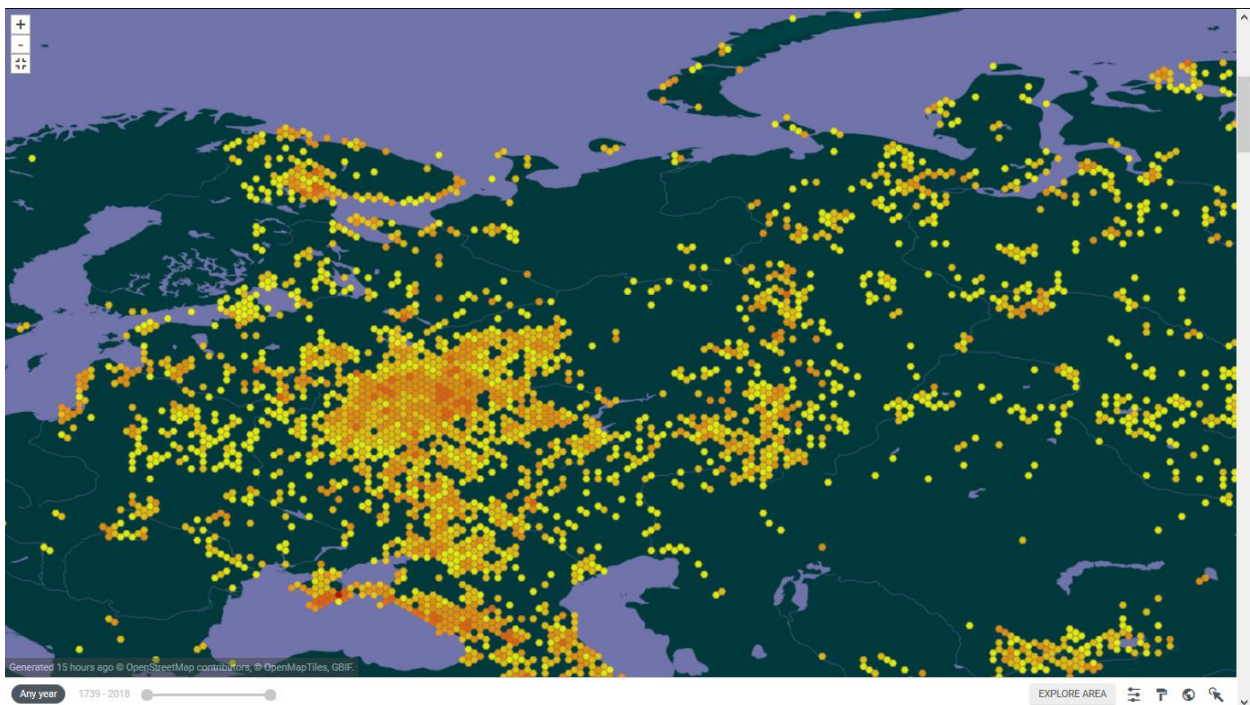


Fig. 1. Eastern Europe and adjacent areas covered by the "Moscow University Herbarium (MW)" dataset (<https://doi.org/10.15468/cpnhcc>) showing 121,960 georeferences within European Russia (as of 01.03.2019).

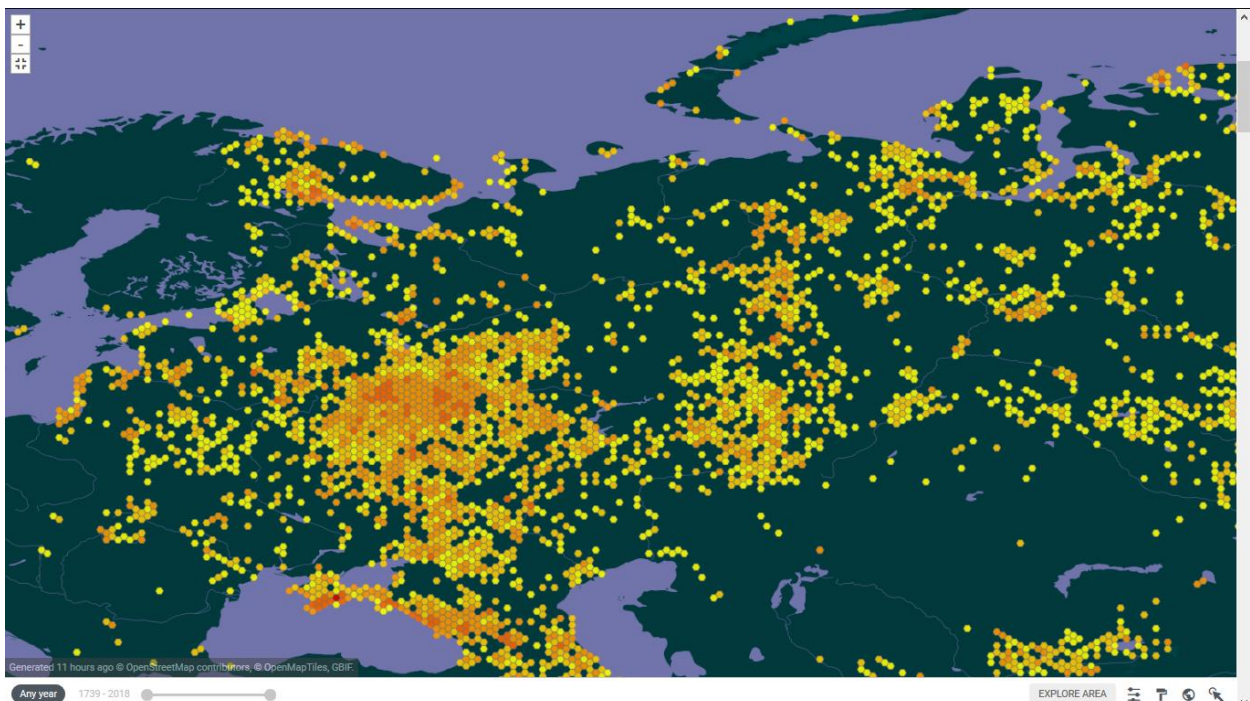


Fig. 2. Eastern Europe and adjacent areas covered by the "Moscow University Herbarium (MW)" dataset (<https://doi.org/10.15468/cpnhcc>) showing 158,189 georeferences within European Russia (as of 15.05.2019).

The difference between two maps is not much impressive due to good sampling of the collection prior to the starting point of the project. Nonetheless, one should mind new plant diversity data available for Komi Republic, denser sampling of western slope of the Middle Urals and Bashkortostan, better data availability in the Moscow area, and new records in Central Chernozem Region.

The spatial coverage of the geodata with a brief description of the spreadsheets integrated both into the Moscow Digital Herbarium (<https://plant.depo.msu.ru/>) and the Moscow University Herbarium GBIF-dataset (<https://doi.org/10.15468/cpnhcc>) is given below. The spatial data were analyzed at 15.05.2019.

Herbarium Curatorial Areas	Code	Georefs Present	No Georefs	Number of Specimens	With Georefs, %
<i>EUROPEAN RUSSIA</i>					
N Region	E1	18,175	18,951	37,126	49
NW Region	E2	5,408	8,160	13,568	40
W Region	E3	2,247	4,447	6,694	34
Central Region	E4	30,625	21,175	51,800	59
Moscow Region	E4a	27,653	46,729	74,382	37
Central Forest Region	E5	11,878	2,813	14,691	81
Central Chernozem Region	E6	14,099	33,429	47,528	30
Volga-Kama Region	E7	2,661	10,700	13,361	20
Middle Volga Region	E8	5,000	16,269	21,269	24
Lower Volga Region	E9	4,376	18,438	22,814	19
E Region	E10	7,155	15,361	22,516	32
Rostov Oblast	E12a	5,984	1,344	7,328	82
<i>NORTHERN CAUCASUS</i>					
Northern Caucasus – W	K1a	6,068	8,634	14,702	41
Northern Caucasus – Central	K1b	7,659	11,360	19,019	40
Northern Caucasus – E	K1c	2,788	5,476	8,264	34
Dagestan	K2	1,984	3,727	5,711	35
Russian Black Sea Coast	K3	4,430	4,484	8,914	50

The following twelve major spreadsheets with geodata incorporated from March, 1 to May, 15 were the most essential:

- 1) Georeferences from the Urals by V.P. Travkin, set #1 (462 georefs);

- 2) Georeferences from Eastern Europe and Northern Caucasus (general) by gbif.ru team (3,387 georefs);
- 3) Georeferences from Eastern Europe for specimens collected after 2000 by N. Kopylova (399 georefs);
- 4) Georeferences from Moscow Region for specimens collected in 1950-1999 by I. Krivokorin (1,032 georefs);
- 5) Georeferences from the City of Moscow (retrieved by OCR) by I. Melnik (453 georefs);
- 6) Georeferences from the Urals by V.P. Travkin, set #2 (1,095 georefs);
- 7) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by I. Krivokorin, set #1 (1,958 georefs);
- 8) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by I. Krivokorin, set #2 (752 georefs);
- 9) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by A.P. Seregin (4,814 georefs);
- 10) Georeferences from Arkhangelsk Oblast and Komi by K.V. Dudova (5,773 georefs);
- 11) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by I. Krivokorin, set #3 (1,685 georefs);
- 12) Text transcriptions of ca. 144,000 labels (made by the commercial partner in 2018) with consequent automatic georeferencing (2,874 georefs).

These twelve spreadsheets include 24,684 georeferences. Other 11,815 newly incorporated georeferences were made by the herbarium staff members as their daily routine job and inserted through the web-interface directly into the Moscow Digital Herbarium (<https://plant.depo.msu.ru/>). All together, we georeferenced 36,499 specimens by May, 15 2019 completing the mission on 67%.

3. Feedback on the project's relevance from the partners and stakeholders

Using the "Activity" bookmark, we detected an intensification of uploading data events from our GBIF-dataset. In late March 2019 the average number of daily uploading events from our dataset was about 46, while in mid-May 2019 it is ca. 52. About 25% of download events employs "Has coordinates" filter, therefore the growing number of georeferences in the Moscow Digital Herbarium leads to the better visibility of our holdings.

The number of citations (as seen in gbif.org) is also rapidly growing. GBIF has established itself as an essential infrastructure underpinning biodiversity-related science. This is clearly demonstrated by the increasing volume of peer-reviewed research using GBIF-mediated data. As of 15.05.2019, our GBIF-dataset accumulated 55 citations—this number doubled since mid-November 2018. However, it is impossible to figure out whether the growing number of citations is one of the project's deliverables or not.

4. Comments on the project implementation, its efficiency and effectiveness; the management arrangements for the project, including support from the GBIF Secretariat

The project is operated not as an isolated management unit, but within deeply cooperated system of data mobilization and management established in the Moscow University Herbarium. The IT-infrastructure already developed in the Moscow University includes a web-portal (<https://plant.depo.msu.ru/>), GBIF IPT with ten published datasets and 1,242,085 records, and an IT-unit with three permanent staff members. This helped us to concentrate our labour and budget efforts on generating of a new content.

Since September 2018, we made several applications for grants across the Russian funding agencies. Unfortunately, we failed to raise budget for the whole process to reach triple-100% for imaging, label capturing and georeferencing. Only some regional proposals of minor importance were supported. They are devoted (1) to the field floristic research in the Moscow metropolitan area poorly represented in our collection (since December 2018, 2 years); (2) study of the spatial patterns of the Krasnodar Krai flora (since April 2019, 2 years); and (3) digitization of the local herbaria in Tula Oblast using depo.msu.ru infrastructure (since April 2019, 3 years).

Therefore, the supporting of the regular georeferencing of the European Russia holdings in the Moscow Digital Herbarium by FinBIF become an important resource for promoting further data mobilization within a larger region which is especially well represented in the collection.

5. Any reflection on the mid-term evaluation itself that could help inform the project's final evaluation and final report; Areas of success to build on during the remainder of the project implementation

We are planning to focus our efforts on annoying spatial gaps present in Northern Region of European Russia, Volga-Kama Region and Lower Volga Region. We do not have many holdings from that regions, but we need more progress in these areas.

Evidently, some new geodata will be mobilized from North-Western Region, Western Region, Central Region, Moscow Region, Central Chernozem Region, Middle Volga Region, Eastern Region, as well as from Northern Caucasus.

We are not expecting new georeferences from Central Forest Region (Kostroma, Yaroslavl, and Ivanovo Oblasts) and Rostov Oblast where more than 80% of holdings have already got coordinates due to special regional programs implemented in the Moscow University Herbarium 2017–2018.

6. The main lessons from the project experience that could be applied in other contexts, including any best practice that others in the GBIF community could apply.

What lessons we learned after the finishing of the Noah's Ark megagrant (2015–2018) which aimed to support the Moscow University biological collections?

(1) Never give up! We applied for 16 grants, competition and prizes to keep our work moving. The most precious trophies were not conquered, but nonetheless we have four modest grants to promote regional missions and a contract with GBIF Secretariat funded by FinBIF.

(2) Do not rely on volunteers in your work. You will spend dozens of hours to teach a volunteer with no guarantee that he will come tomorrow. Think twice, do you really need him? Only staff members of the Moscow University Herbarium and freelancers permanently involved in our daily activities are involved in the project.

(3) Virtual consolidation of lesser collections within a larger web-hub is a budget-efficient solution. We mobilized ca. 16,000 herbarium specimens from the Main Botanical Garden of RAS (MHA) within "Moscow project". Herbarium of the Main Botanical Garden is the second largest in Moscow, and this will be their first appearance in GBIF as soon as they will be registered as a publisher. Their dataset will be published via the Moscow University IPT.

7. Changes to the project plans: a disclaimer.

The plan of the project submitted to GBIF Secretariat as part of the application form and later included to the contract did not include any unrealistic intensions. Therefore, we do not expect any substantial changes in the project plans. We are planning to finish

georeferencing by mid-September 2019, after that to check 26,000 label transcriptions made by the commercial partner and to synchronize geodata and text data on the portal within one or two working days. The final report will be submitted to GBIF Secretariat by September, 30 2019.

Annex – Sources of verification

1. Links to videos with talks by Dr. A.P. Seregin:

<https://www.youtube.com/watch?v=n3GjVwyDGtQ> (from 15:00) – Digitization and mobilization of the biodiversity data in the Moscow University

<https://www.youtube.com/watch?v=xkZISUxf7M8> – An introduction: Why GBIF is important for us?

<https://www.youtube.com/watch?v=ZC52zYIsUn0> – Data cleaning for advanced users

2. Links to the posts in the Herbarium blog:

https://vk.com/wall-134484155_1932 – Russia in GBIF – a brief review (8.03.2019)

https://vk.com/wall-134484155_1936 – new geodata in the Moscow Digital Herbarium (9.03.2019)

https://vk.com/wall-134484155_1943 – new geodata in the Moscow Digital Herbarium (12.03.2019)

https://vk.com/wall-134484155_1966 – new geodata in the Moscow Digital Herbarium (18.03.2019)

https://vk.com/wall-134484155_1976 – new geodata in the Moscow Digital Herbarium (20.03.2019)

https://vk.com/wall-134484155_1996 – new text data in the Moscow Digital Herbarium (24.03.2019)

https://vk.com/wall-134484155_2014 – new geodata in the Moscow Digital Herbarium (29.03.2019)

https://vk.com/wall-134484155_2021 – new geodata in the Moscow Digital Herbarium (30.03.2019)

https://vk.com/wall-134484155_2022 – new geodata in the Moscow Digital Herbarium (31.03.2019)

https://vk.com/wall-134484155_2085 – new dataset from the Moscow University on GBIF
(18.04.2019)